Chapter 6

Layer 3 VPN Overview

The JUNOS software implements Layer 3 BGP/MPLS virtual private networks (VPNs) as defined in RFC 2547 and Internet draft draft-rosen-rfc2547bis (also referred to as RFC 2547bis). This chapter discusses the following topics that provide background information about Layer 3 VPNs:

Layer 3 VPN Overview

In JUNOS, Layer 3 VPNs are based on RFC 2547bis. RFC 2547bis defines a mechanism by which service providers can use their IP backbones to provide VPN services to their customers. A Layer 3 VPN is a set of sites that share common routing information and whose connectivity is controlled by a collection of policies. The sites that make up a Layer 3 VPN are connected over a provider's existing public Internet backbone.

RFC 2547bis VPNs are also known as BGP/MPLS VPNs because BGP is used to distribute VPN routing information across the provider's backbone, and MPLS is used to forward VPN traffic across the backbone to remote VPN sites.

Customer networks, because they are private, can use either public addresses or private addresses, as defined in RFC 1918. When customer networks that use private addresses connect to the public Internet infrastructure, the private addresses might overlap with the same private addresses used by other network users. MPLS/BGP VPNs solve this problem by prefixing a VPN identifier to each address from a particular VPN site, thereby creating an address that is unique both within the VPN and within the public Internet. In addition, each VPN has its own VPN-specific routing table that contains the routing information for that VPN only.

Layer 3 VPN Standards

Layer 3 VPNs are defined in the following documents:

RFC 2547, *BGP/MPLS VPNs*

BGP/MPLS VPNs, Internet draft draft-rosen-rfc2547bis

RFC 2283, *Multiprotocol Extensions for BGP4*

To access Internet RFCs and drafts, go to the IETF Web site at http://www.ietf.org.

Layer 3 VPN Attributes

Route distribution within a VPN is controlled using BGP extended community attributes. RFC 2547 defines the following three attributes used by VPNs:

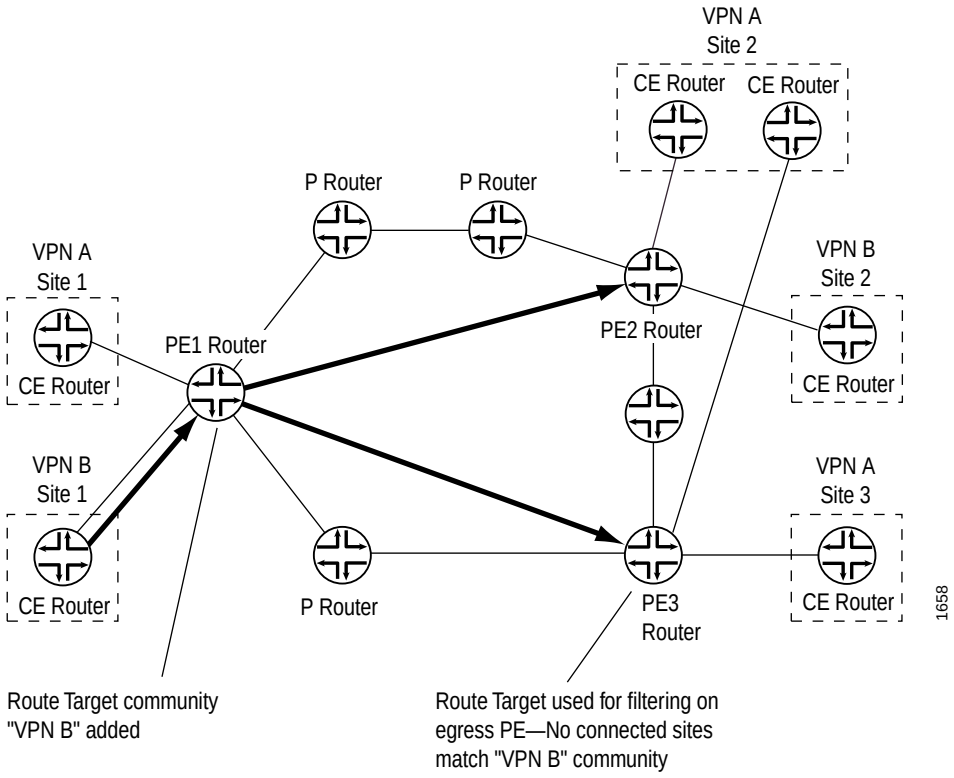
Target VPN—Identifies a set of sites within a VPN to which a PE router distributes routes. This attribute is also called the *route target*. The route target is used by the egress PE router to determine whether a received route is destined for a VPN that the router services.

Figure 4 illustrates the function of the route target. PE router PE1 adds the route target "VPN B" to routes received from the CE router at Site 1 in VPN B. When it receives the route, the egress router PE2 examines the route target, determines that the route is for a VPN that it services, and accepts the route. When the egress router PE3 receives the same route, it does not accept the route because it does not service any CE routers in VPN B.

VPN of origin—Identifies a set of sites and the corresponding route as having come from one of the sites in that set.

Site of origin—Uniquely identifies the set of routes that a PE router learned from a particular site. This attribute ensures that a route learned from a particular site through a particular PE-CE connection is not distributed back to the site through a different PE-CE connection. It is particularly useful if you are using BGP as the routing protocol between the PE and CE routers and if different sites in the VPN have not been assigned distinct AS numbers.

Figure 4: VPN Attributes and Route Distribution

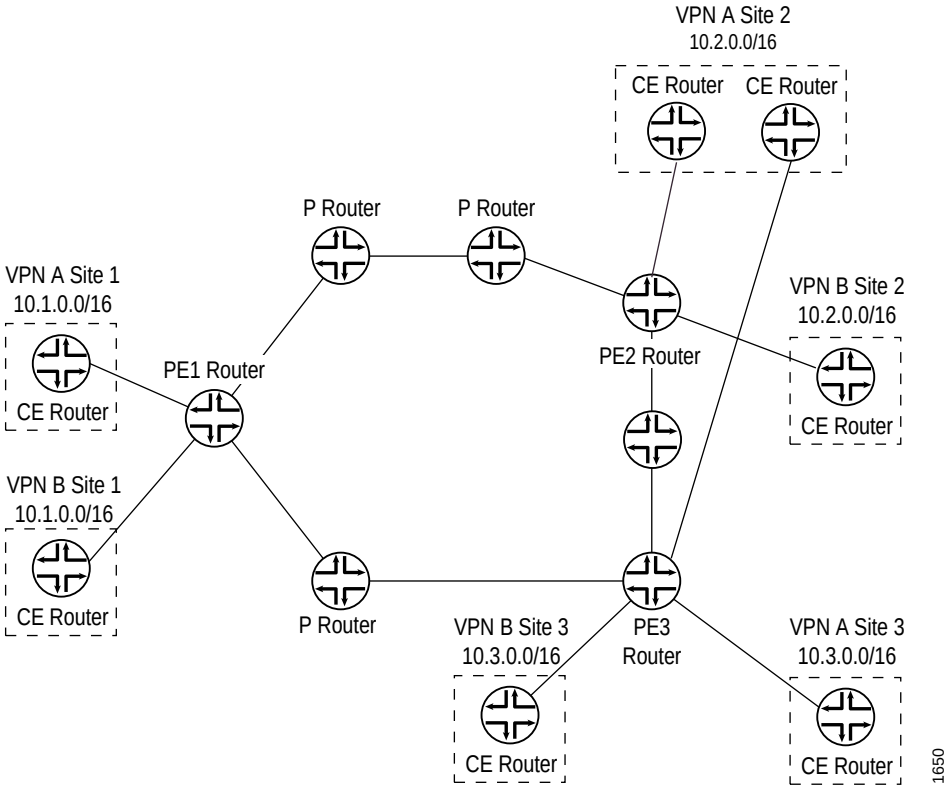


VPN-IPv4 Addresses and Route Distinguishers

Because Layer 3 VPNs connect private networks—which can use either public addresses or private addresses, as defined in RFC 1918—over the public Internet infrastructure, when the private networks use private addresses, the addresses might overlap with the addresses of another private network.

Figure 6 illustrates how private addresses of different private networks can overlap. Here, sites within VPN A and VPN B use the address spaces 10.1.0.0/16, 10.2.0.0/16, and 10.3.0.0/16 for their private networks.

Figure 5: Overlapping Addresses among Different VPNs



To avoid overlapping private addresses, you can configure the network devices to use public addresses instead of private addresses. However, this is a large and complex undertaking. The solution provided in RFC 2547bis uses the existing private network numbers to create a new address that is unambiguous. The new address is part of the VPN-IPv4 address family, which is a BGP address family added as an extension to the BGP protocol. In VPN-IPv4 addresses, a value that identifies the VPN, called a route distinguisher, is prefixed to the private IPv4 address, providing an address that uniquely identifies a private IPv4 address.

Only the PE routers need to support the VPN-IPv4 address extension to BGP. When an ingress PE router receives an IPv4 route from a device within a VPN, it converts it into a VPN-IPv4 route by prefixing the route distinguisher to the route. The VPN-IPv4 addresses are used only for routes exchanged between PE routers. When an egress PE router receives a VPN-IPv4 route, it converts it back to an IPv4 route, by removing the route distinguisher, before announcing the route to its connected CE routers.

VPN-IPv4 addresses have the following format:

Route distinguisher is a 6-byte value that you can specify in one of the following formats:

as-number:*number*, where *as-number* is an AS number (a 2-byte value) and *number* is any 4-byte value. The AS number can be in the range 1 through 65,535. We recommend that you use an IANA-assigned, nonprivate AS number, preferably the ISP's own or the customer's own AS number.

ip-address:*number*, where *ip-address* is an IP address (a 4-byte value) and *number* is any 2-byte value. The IP address can be any globally unique unicast address. We recommend that you use the address that you configure in the router-id statement, which is a nonprivate address in your assigned prefix range.

IPv4 address—4-byte address of a device within the VPN.

Figure 5 illustrates how the AS number can be used in the route distinguisher. Suppose that VPN A is in AS 65535 and that VPN B is in AS 666 (both these AS numbers belong to the ISP), and suppose that the route distinguisher for Site 2 in VPN A is 65535:02 and that the route distinguisher for Site 2 in VPN B is 666:01. When Router PE2 receives a route from the CE router in VPN A, it converts it from its IP address of 10.2.0.0 to a VPN-IPv4 address of 65535:02:10.2.0.0. When the PE router receives a route from VPN B, which uses the same address space as VPN A, it converts it to a VPN-IPv4 address of 666:02:10.2.0.0.

If the IP address is used in the route distinguisher, suppose the Router PE2's IP address is 172.168.0.1. When the PE router receives a route from VPN A, it converts it to a VPN-IPv4 address of 172.168.0.1:0:10.2.0.0/16, and it converts a route from VPN B to 172.168.0.0:1:10.2.0.0/16.

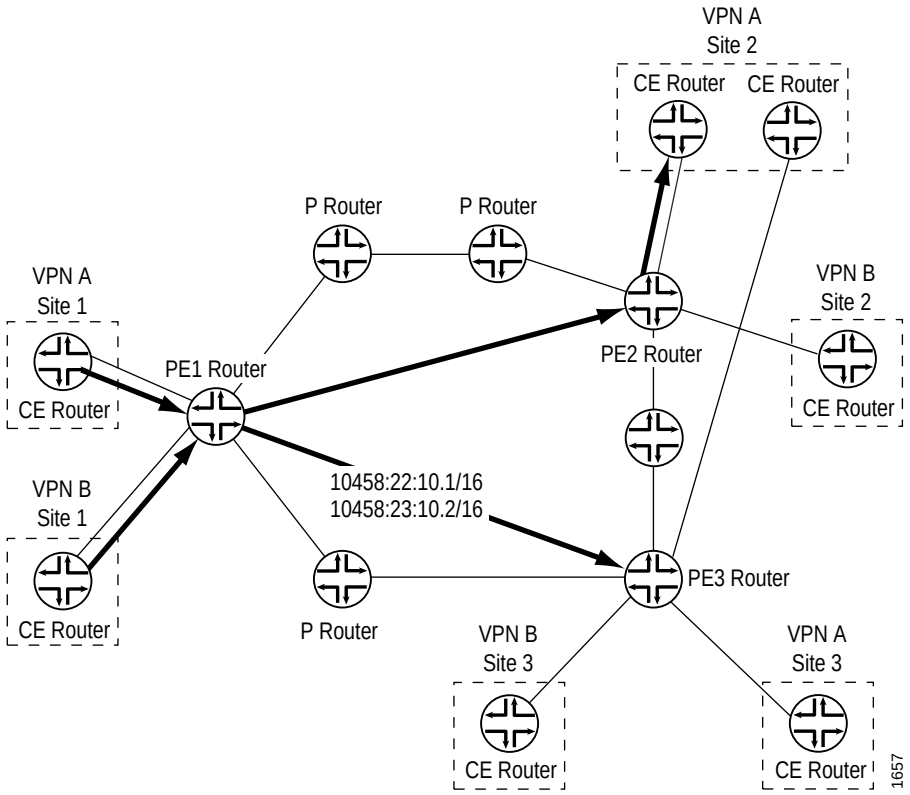
Route distinguishers are used only among PE routers to disambiguate IPv4 addresses from different VPNs. The ingress PE router creates a route distinguisher and converts IPv4 routes received from CE routers into VPN-IPv4 addresses. The egress PE routers convert VPN-IPv4 routes into IPv4 routes before announcing them to the CE router.

Because VPN-IPv4 addresses are a type of BGP address, you must configure IBGP sessions between pairs of PE routers so that the PE routers can distribute VPN-IPv4 routes within the provider's core network. (All PE routers are assumed to be within the same AS.)

You define BGP communities to constrain the distribution of routes among the PE routers. Defining BGP communities does not, by itself, disambiguate IPv4 addresses.

Figure 6 illustrates how Router PE1 adds the route distinguisher 10458:22:10.1/16 to routes received from the CE router at Site 1 in VPN A and forwards these routes to the other two PE routers. Similarly, Router PE1 adds the route distinguisher 10458:23:10.2/16 to routes received by the CE router at Site 1 in VPN B and forwards these routes to the other PE routers.

Figure 6: Route Distinguishers

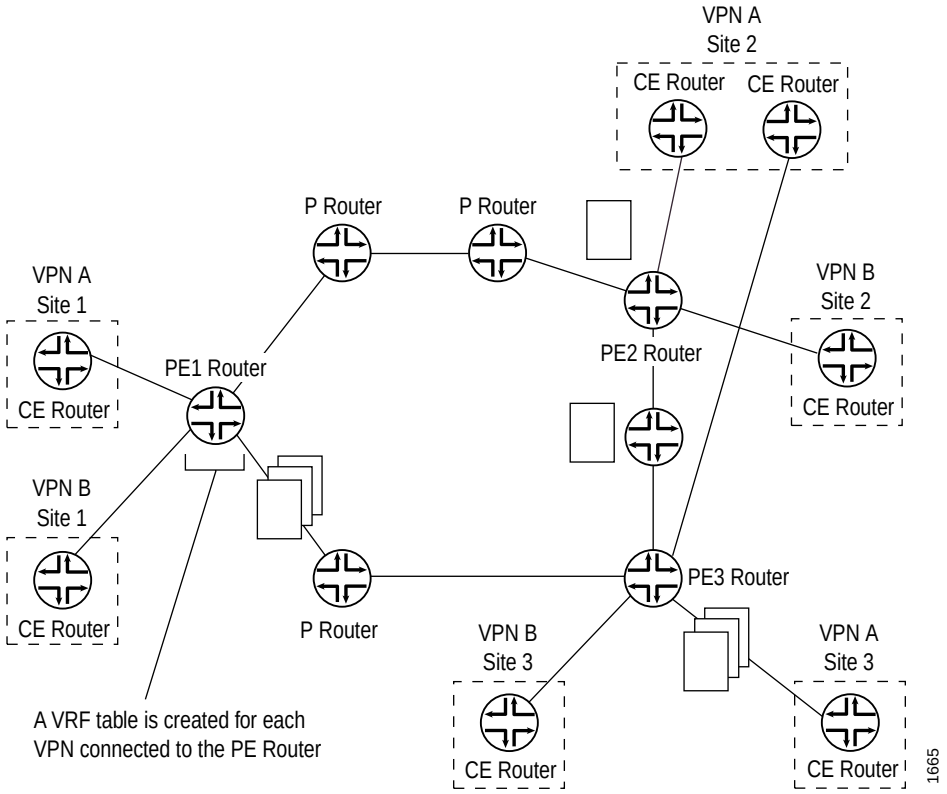


VPN Routing and Forwarding Tables

To separate a VPN's routes from routes in the public Internet or those in other VPNs, the PE router creates a separate routing table for each VPN, called a VPN Routing and Forwarding (VRF) table. The PE router creates one VRF table for each VPN that has a connection to a CE router. Any customer or site that belongs to the VPN can access only the routes in the VRF tables for that VPN.

Figure 7 illustrates the VRF tables that are created on the PE routers. The three PE routers have connections to CE routers that are in two different VPNs, so each of these PE routers creates two VRF tables, one for each VPN.

Figure 7: VRF Tables



Each VRF table is populated from routes received from directly connected CE sites associated with that VRF and from routes received from other PE routers that passed BGP community filtering and are in the same VPN.

Each PE router also maintains one global routing table (inet.0) to reach other routers in and outside the provider's core network.

Each customer connection (that is, each logical interface) is associated with one VRF table. Only the VRF table associated with a customer site is consulted for packets from that site.

You can configure the router so that if a next hop to a destination is not found in the VRF table, the router performs a lookup in the global routing table, which is used for Internet access.

The JUNOS software uses the following routing tables for VPNs:

bgp.l3vpn.0—Stores all VPN-IPv4 unicast routes received from other PE routers. (This table does not store routes received from directly connected CE routers.) This table is present only on PE routers.

When a PE router receives a route from another PE router, it places the route into its bgp.l3vpn.0 routing table. The route is resolved using the information in the inet.3 routing table. The resultant route is converted into IPv4 format and redistributed to all *routing-instance-name*.inet.0 routing tables on the PE router if it matches the VRF import policy.

The bgp.l3vpn.0 table is also used to resolve routes over the MPLS tunnels that connect the PE routers. These routes are stored in the inet.3 routing table. PE-PE router connectivity must exist in inet.3 (not just in inet.0) for VPN routes to be resolved properly.

To determine whether to add a route to the bgp.l3vpn.0 routing table, the JUNOS software checks it against the VRF import policies for all the VPNs configured on the PE router. If the VPN-IPv4 route matches one of the policies, it is added to the bgp.l3vpn.0 table. To display the routes in the bgp.l3vpn.0 routing table, use the show route table bgp.l3vpn.0 command.

routing-instance-name.inet.0—Stores all unicast IPv4 routes received from directly connected CE routers in a routing instance (that is, in a single VPN) and all explicitly configured static routes in the routing instance. This is the VRF table and is present only on PE routers. For example, for a routing instance named VPN-A, the routing table for that instance is named VPN-A.inet.0.

When a CE router advertises to a PE router, the PE router places the route into the corresponding *routing-instance-name*.inet.0 routing table and advertises the route to other PE routers if it passes a VRF export policy. Among other things, this policy tags the route with the route distinguisher (route target) that corresponds to the VPN site to which the CE belongs. A label is also allocated and distributed with the route. The bgp.l3vpn.0 routing table is not involved in this process.

The *routing-instance-name*.inet.0 table also stores routes announced by a remote PE router that match the VRF import policy for that VPN. The remote PE router redistributed these routes from its bgp.l3vpn.0 table.

Routes are not redistributed from the *routing-instance-name*.inet.0 table to the bgp.l3vpn.0 table; they are directly advertised to other PE routers.

For each *routing-instance-name*.inet.0 routing table, one forwarding table is maintained in the router's Packet Forwarding Engine. This table is maintained in addition to the forwarding tables that correspond to the router's inet.0 and mpls.0 routing tables. As with the inet.0 and mpls.0 routing tables, the best routes from the *routing-instance-name*.inet.0 routing table are placed into the forwarding table.

To display the routes in the *routing-instance-name*.inet.0 table, use the show route table *routing-instance-name*.inet.0 command.

inet.3—Stores all MPLS routes learned from LDP and RSVP signaling done for VPN traffic. The routing table stores the MPLS routes only if the traffic-engineering bgp-igp option is not enabled.

For VPN routes to be resolved properly, the inet.3 table must contain routes to all the PE routers in the VPN.

To display the routes in the inet.3 table, use the show route inet.3 command.

Note that IGP shortcuts do not work in VPN environments and should not be configured. IGP shortcuts move routes in inet.3 to inet.0. VPN IBGP (family inet-vpn) relies on next-hops that are in the inet.3 table; thus, IGP shortcuts are incompatible with VPNs.

inet.0—Stores routes learned by the IBGP sessions between the PE routers. To provide Internet access to the VPN sites, configure the *routing-instance-name*.inet.0 routing table to contain a default route to the inet.0 routing table.

To display the routes in the inet.0 table, use the show route inet.0 command.

The following routing policies, which are defined in VRF import and export statements, are specific to VRF tables.

Import policy—Applied to VPN-IPv4 routes learned from another PE router to determine whether the route should be added to the PE router's bgp.l3vpn.0 routing table. Each routing instance on a PE router has a VRF import policy.

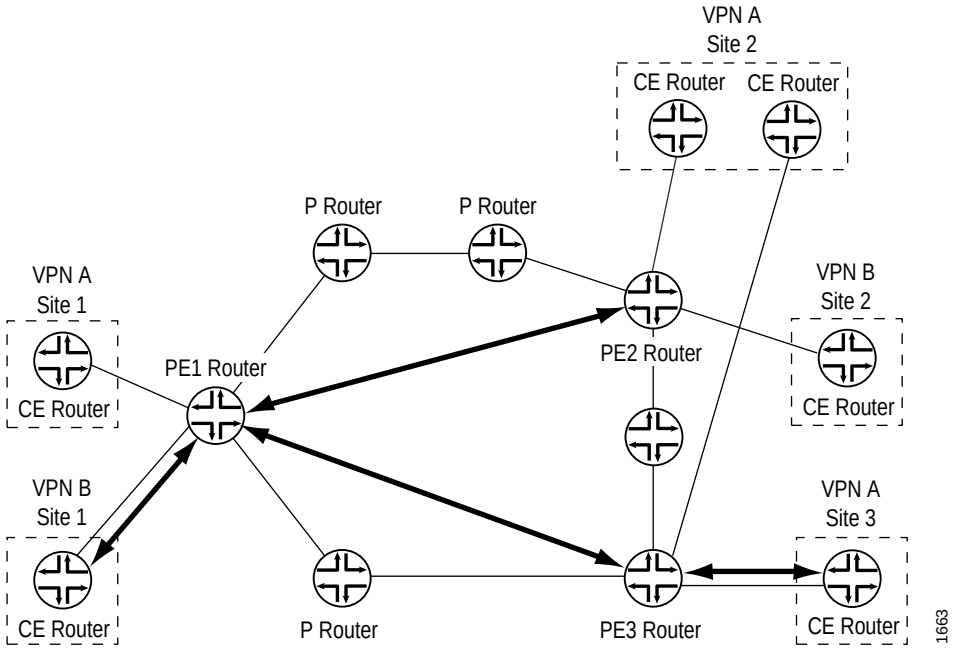
Export policy—Applied to VPN-IPv4 routes that are announced to other PE routers. The VPN-IPv4 routes are IPv4 routes that have been announced by locally connected CE routers.

VPN route processing differs from normal BGP route processing in one way. In BGP, routes are accepted if they are not explicitly rejected by import policy. However, because many more VPN routes are expected, the JUNOS software does not accept (and hence store) VPN routes unless the route matches at least one VRF import policy. If no VRF import explicitly accepts the route, it is discarded and not even stored in the bgp.l3vpn.0 table. As a result, if a VPN change occurs on a PE router—such as adding a new VRF table or changing a VRF import policy—the PE router sends a BGP route refresh message to the other PE routers (or to the route reflector if this is part of the VPN topology) to retrieve all VPN routes so they can be re-evaluated to determine whether they should be kept or discarded.

Route Distribution within a Layer 3 VPN

Within a VPN, the distribution of VPN-IPv4 routes occurs between the PE and CE routers and between the PE routers (see Figure 8).

Figure 8: Route Distribution within a VPN



This section discusses the following:

Distribution of Routes from CE to PE Routers

A CE router announces its routes to the directly connected PE router. The announced routes are in IPv4 format. The PE router places the routes into the VRF table for the VPN. In the JUNOS software, this is the *routing-instance-name*.inet.0 routing table, where *routing-instance-name* is the configured name of the VPN.

The connection between the CE and PE routers can be a remote connection (a WAN connection) or a direct connection (such as a Frame Relay or Ethernet connection).

CE routers can communicate with PE routers using one of the following:

OSPF

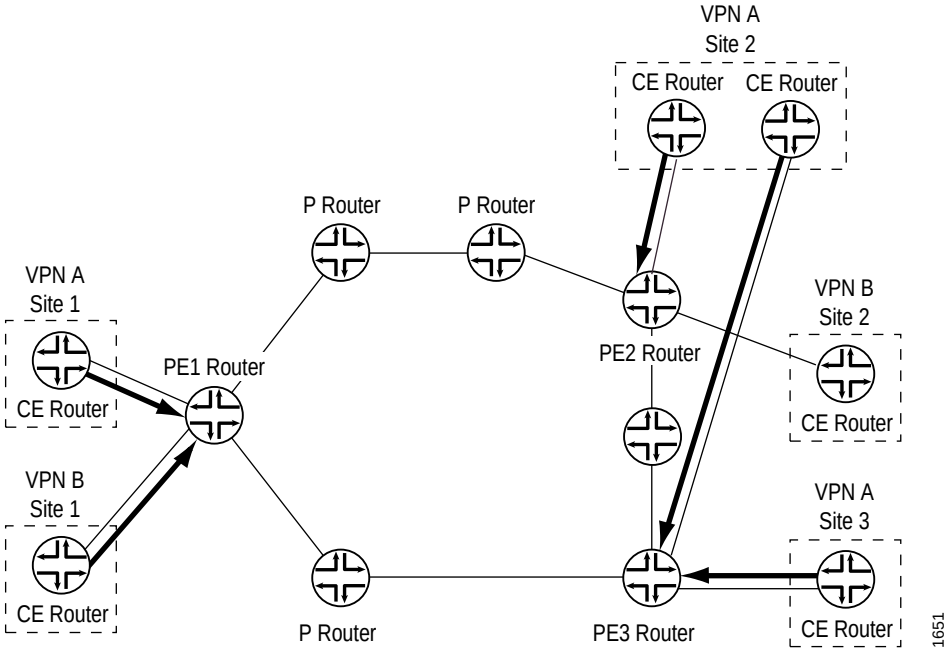
RIP

BGP

Static route

Figure 9 illustrates how routes are distributed from CE routers to PE routers. Router PE1 is connected to two CE routers that are in different VPNs. Therefore, it creates two VRF tables, one for each VPN. The CE routers announce IPv4 routes. The PE router installs these routes into two different VRF tables, one for each VPN. Similarly, Router PE2 creates two VRF tables into which routes are installed from the two directly connected CE routers. Router PE3 creates one VRF table because it is directly connected to only one VPN.

Figure 9: Distribution of Routes from CE Routers to PE Routers



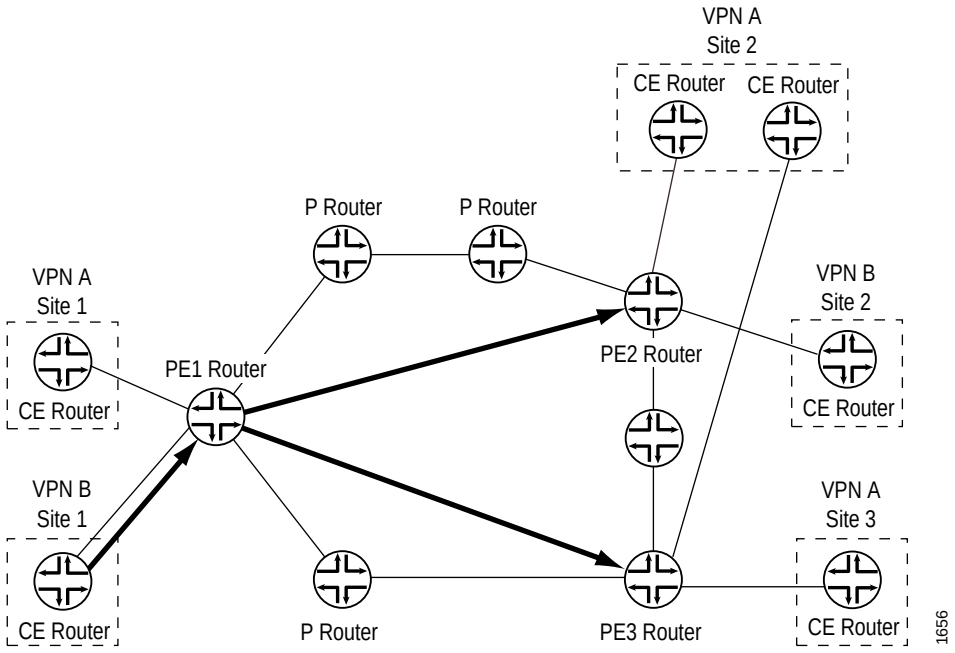
Distribution of Routes between PE Routers

When one PE router receives routes advertised from a directly connected CE router, it checks the received route against the VRF export policy for that VPN. If it matches, the route is converted to VPN-IPv4 format—that is, the route distinguisher (route target) is added to the route. The PE router then announces the route in VPN-IPv4 format to the remote PE routers. The routes are distributed using IBGP sessions, which are configured in the provider's core network. If the route does not match, it is not exported to other PE routers, but can still be used locally for routing, for example, if two CE routers in the same VPN are directly connected to the same PE router.

The remote PE router places the route into its bgp.l3vpn.0 table if the route passes the import policy on the IBGP session between the PE routers. At the same time, it checks the route against the VRF import policy for the VPN. If it matches, the route distinguisher is removed from the route and it is placed into the VRF table (the *routing-instance-name*.inet.0 table) in IPv4 format.

Figure 10 illustrates how Router PE1 distributes routes to the other PE routers in the provider's core network. Router PE2 and Router PE3 each have VRF import policies that they use to determine whether to accept routes received over the IBGP sessions and install them in their VRF tables.

Figure 10: Distribution of Routes between PE Routers



Distribution of Routes from PE to CE Routers

The remote PE router announces the routes in its VRF tables, which are in IPv4 format, to its directly connected CE routers.

PE routers can communicate with CE routers using one of the following routing protocols:

OSPF

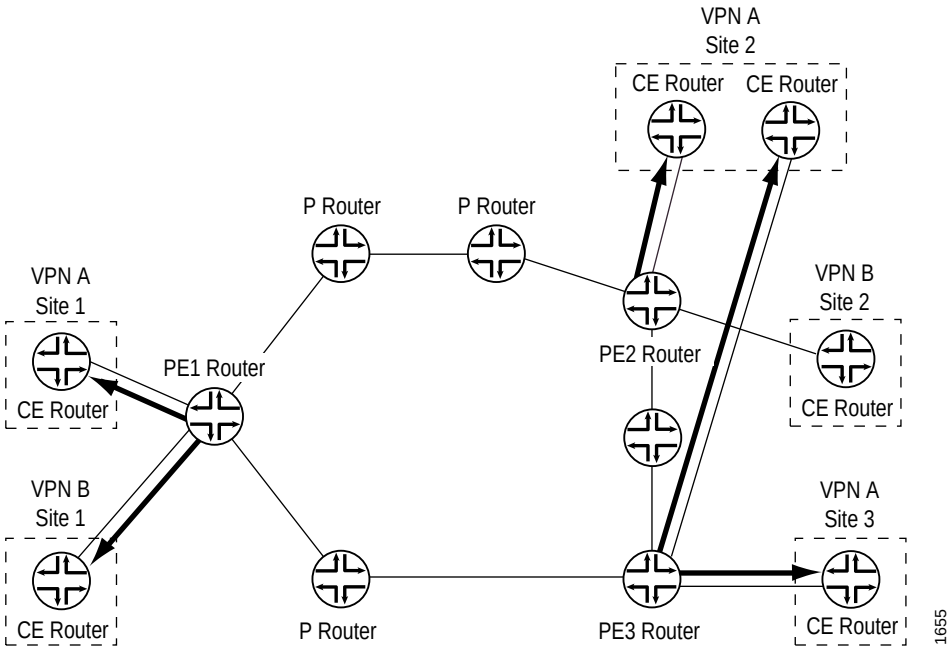
RIP

BGP

Static route

Figure 11 illustrates how the three PE routers announce their routes to their connected CE routers.

Figure 11: Distribution of Routes from PE Routers to CE Routers



Forwarding across the Provider's Core Network

The PE routers in the provider's core network are the only routers that are configured to support VPNs and hence are the only routers that know about the existence of the VPNs. From the point of view of VPN functionality, the provider routers in the core—those provider routers that are not directly connected to CE routers—are merely routers along the tunnel between the ingress and egress PE routers.

The tunnels can be either LDP or MPLS. Any provider routers along the tunnel must support the protocol used for the tunnel, either LDP or MPLS.

When PE-router-to-PE router forwarding is tunneled over MPLS LSPs, the MPLS packets have a two-level label stack (see Figure 12):

Outer label—Label assigned to the address of the BGP next hop by the IGP next hop

Inner label—Label that the BGP next hop assigned for the packet's destination address

Figure 12: Using MPLS LSPs to Tunnel between PE Routers

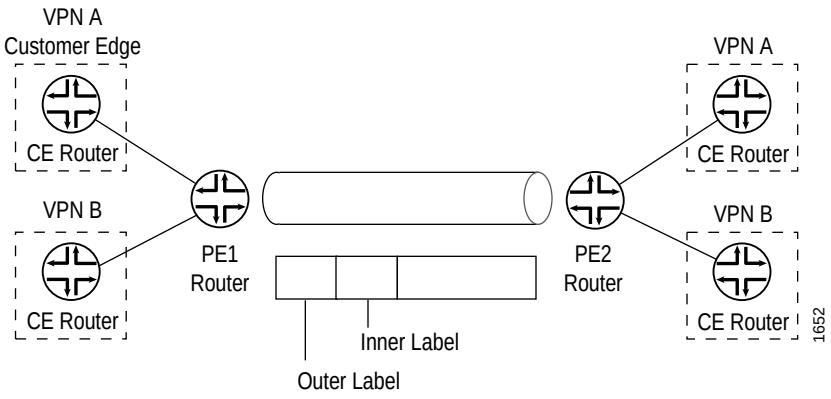


Figure 13 illustrates how the labels are assigned and removed:

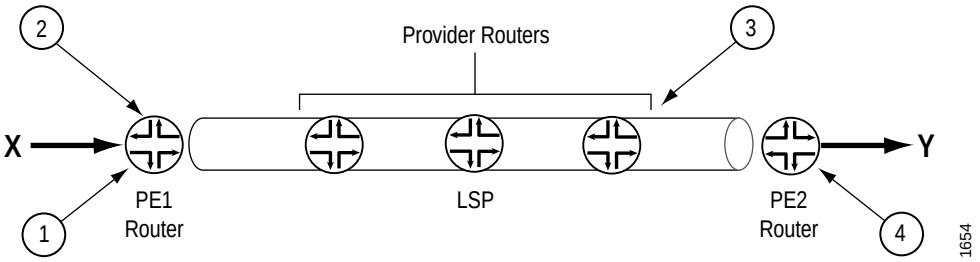
1. When CE Router X forwards a packet to Router PE1 with a destination of CE Router Y, the PE route identifies the BGP next hop to Router Y and assigns a label that corresponds to the BGP next hop and identifies the destination CE router. This label is the inner label.

2. Router PE1 then identifies the IGP route to the BGP next hop and assigns a second label that corresponds to the LSP of the BGP next hop. This label is the outer label.

3. The inner label remains the same as the packet traverses the LSP tunnel. The outer label is swapped at each hop along the LSP and is then popped by the penultimate hop router (the third provider router).

4. Router PE2 pops the inner label from the route and forwards the packet to Router Y.

Figure 13: Label Stack



Routing Instances for VPNs

To implement Layer 3 VPNs in the JUNOS software, you configure one routing instance for each VPN. You configure the routing instances on PE routers only. Each VPN routing instance consists of the following components:

VRF table—On each PE router, you configure one VRF table for each VPN.

Set of interfaces that use the VRF table—The logical interface to each directly connected CE router must be associated with a VRF table. You can associate more than one interface with the same VRF table if more than one CE router in a VPN is directly connected to the PE router.

Policy rules—These control the import of routes into and the export of routes from the VRF table.

One or more routing protocols that install routes from CE routers into the VRF table—You can use the BGP, OSPF, and RIP routing protocols, and you can use static routes.

Multicast Over Layer 3 VPNs

You can configure multicast routing over a network running a Layer 3 VPN that complies with RFC 2547. This section describes this type of network application, and includes these topics:

Multicast Over Layer 3 VPNs Overview on page 66

Sending PIM Hello Messages to the PE Routers on page 67

Sending PIM Join Messages to the PE Routers on page 68

Receiving the Multicast Transmission on page 68

Multicast Over Layer 3 VPNs Overview

In the unicast environment of a Layer 3 VPN, all VPN state information is contained within the PE routers. In a multicast Layer 3 VPN environment, Protocol Independent Multicast (PIM) adjacencies are established between the CE router and the PE router and between the master PIM instance. They are configured at the [protocols pim] hierarchy level on the IGP neighbors of the PE router. The set of master PIM adjacencies on the service provider's network make up the forwarding path, which consists of a rendezvous point (RP) tree rooted at the RP within the service provider's network.

Therefore, provider (P) routers within the provider network must maintain multicast state information for the Layer 3 VPNs. For this to function, there must be two types of rendezvous points for each VPN:

- The VPN-RP, an RP that resides within the VPN
- The service provider RP (SP-RP), which resides within the service provider network

A PE router can act as an SP-RP, but cannot be the VPN-RP of a Layer 3 VPN. The VPN-RP must be located on a CE router or some other customer router within the VPN.

To configure multicast over a Layer 3 VPN, you must install a Tunnel Services PIC on the following devices:

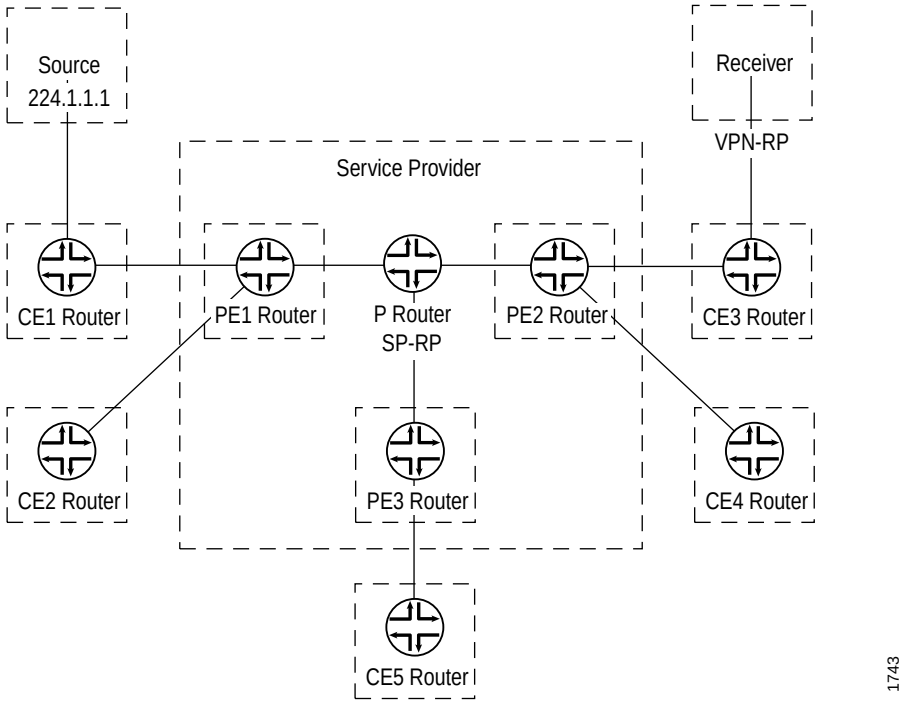
- Provider routers acting as rendezvous points
- PE routers configured to run multicast routing
- CE routers acting as destination routers or as VPN-RPs

For more information about running multicast over Layer 3 VPNs, see the following documents:

- *Multicast in MPLS/BGP VPNs* , Internet draft draft-rosen-vpn-mcast-02.txt
- *JUNOS Internet Softw are Configur ation Guide: Multicast*

The sections that follow describe the operation of a multicast VPN. Figure 14 illustrates the network topology used.

Figure 14: Multicast Topology Overview



Sending PIM Hello Messages to the PE Routers

The first step in initializing multicast over a Layer 3 VPN is the distribution of a PIM Hello message from a PE router (called PE3 in this section) to all the other PE routers on which PIM is configured.

You configure PIM on the Layer 3 VPN routing instance on the PE3 router. If a Tunnel Services PIC exists on the router, a multicast interface is created. This interface is used to communicate between the PIM instance within the VRF and the master PIM instance.

The following occurs when a PIM Hello message is sent to the PE routers:

1. A PIM Hello message is sent from the VRF over the multicast interface. A GRE header is prepended to the PIM Hello message. The header message includes the VPN group address and the loopback address of the PE3 router.

2. A PIM register header is prepended to the Hello message as the packet is looped through the PIM encapsulation interface. This header contains the destination address of the SP-RP and the loopback address of the PE3 router.

3. The packet is sent to the SP-RP.

4. The SP-RP removes the top header from the packet and sends the remaining GRE-encapsulated Hello message to all the PE routers.

5. The master PIM instance on each PE router handles the GRE encapsulated packet. Because the VPN group address is contained in the packet, the master instance removes the GRE header from the packet and sends the Hello message, which contains the proper VPN group address within the VRF, over the multicast interface.

Sending PIM Join Messages to the PE Routers

To receive a multicast broadcast from a multicast network, a CE router must send a PIM Join message to the VPN-RP. The process described in this section refers to Figure 14.

The CE5 router needs to receive a multicast broadcast from multicast source 224.1.1.1. To receive the broadcast, it sends a PIM Join message to the VPN-RP (the PE3 router):

1. The PIM Join message is sent through the multicast interface, and a GRE header is prepended to the message. The GRE header contains the VPN group ID and the loopback address of the PE3 router.

2. The PIM Join message is then sent through the PIM encapsulation interface; a register header is prepended to the packet. The register header contains the IP address of the SP-RP and the loopback address of the PE3 router.

3. The PIM Join message is sent to the SP-RP by means of unicast routing.

4. On the SP-RP, the register header is stripped off (the GRE header remains) and the packet is sent to all the PE routers.

5. The PE2 router receives the packet, and because the link to the VPN-RP is through the PE2 router, it sends the packet through the multicast interface to remove the GRE header.

6. Finally, the PIM Join message is sent to the VPN-RP.

Receiving the Multicast Transmission

The steps that follow outline how a multicast transmission is propagated across the network:

1. The multicast source connected to the CE1 router sends the packet to group 224.1.1.1 (the VPN group address). The packet is encapsulated into a PIM register.

2. Because this packet already includes the PIM header, it is forwarded by means of unicast routing to the VPN-RP over the Layer 3 VPN.

3. The VPN-RP removes the packet and sends it out the downstream interfaces (which include the interface back to the CE3 router). The CE3 router also forwards this to the PE3 router.

4. The packet is sent through the multicast interface on the PE2 router; in the process, the GRE header is prepended to the packet.

5. Next, the packet is sent through the PIM encapsulation interface, where the register header is prepended to the data packet.

6. The packet is then forwarded to the SP-RP, which removes the register header, leaves the GRE header intact, and sends the packet to the PE routers.

7. PE routers remove the GRE header and forward the packet to the CE routers that requested the multicast broadcast by sending the PIM Join message.

PE routers that have not received requests for multicast broadcasts from their connected CE routers still receive packets for the broadcast. These PE routers drop the packets as they are received.